Received 7 August 2023; revised 8 November 2023; accepted 8 November 2023. Date of publication 12 March 2024; date of current version 18 June 2024. The review of this article was arranged by Associate Editor Selin Aviyente.

Digital Object Identifier 10.1109/OJSP.2024.3376300

Person Identification and Relapse Detection From Continuous Recordings of Biosignals Challenge: Overview and Results

ATHANASIA ZLATINTSI, PANAGIOTIS P. FILNTISIS, NIKI EFTHYMIOU, CHRISTOS GAROUFIS, GEORGE RETSINAS, THOMAS SOUNAPOGLOU, ILIAS MAGLOGIANNIS, PANAYIOTIS TSANAKAS, NIKOLAOS SMYRNIS, AND PETROS MARAGOS

[1]School of ECE, National Technical University of Athens, 15773 Athens, Greece
[2]BLOCKACHAIN PC, 55535 Thessaloniki, Greece
[3]Department of Digital Systems, University of Piraeus, 18534 Pireas, Greece
[4]Laboratory of Cognitive Neuroscience and Sensorimotor Control, University Mental Health, Neurosciences and Precision Medicine Research Institute COSTAS STEFANIS, 11527 Athens, Greece
[5]2nd Department of Psychiatry, University General Hospital "ATTIKON," Medical School, National and Kapodistrian University of Athens, 12462 Athens, Greece

CORRESPONDING AUTHOR: ATHANASIA ZLATINTSI (e-mail: nzlat@cs.ntua.gr).

This work was supported by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project acronym: e-Prevention, code: T1EDK-02890/MIS: 5032797).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the General Regulation (EU) 2016/679, and the protocol was approved by the Ethics Committee of the University Mental Health, Neurosciences and Precision Medicine Research Institute "Costas Stefanis" (UMHRI) in Athens, Greece under Application No. T1EDK-02890 / MIS: 5032797. [Online]. Available: https://eprevention.gr/.

ABSTRACT This paper presents an overview of the e-Prevention: Person Identification and Relapse Detection Challenge, which was an open call for researchers at ICASSP-2023. The challenge aimed at the analysis and processing of long-term continuous recordings of biosignals recorded from wearable sensors, namely accelerometers, gyroscopes and heart rate monitors embedded in smartwatches, as well as sleep information and daily step counts, in order to extract high-level representations of the wearer's activity and behavior, termed as digital phenotypes. Specifically, with the goal of analyzing the ability of these digital phenotypes in quantifying behavioral patterns, two tasks were evaluated in two distinct tracks: 1) Identification of the wearer of the smartwatch, and 2) Detection of psychotic relapses in patients in the psychotic spectrum. The long-term data that have been used in this challenge have been acquired during the course of the e-Prevention project (Zlatintsi et al., 2022), an innovative integrated system for medical support that facilitates effective monitoring and relapse prevention in patients with mental disorders. Two baseline systems, one for each task, were described and the validation scores for both tasks were provided to the participants. Herein, we present an overview of the approaches and methods as well as the performance analysis and the results of the 5-top ranked participating teams, which in track 1 achieved accuracy results between 91%-95%, while in track 2 mean PR- and ROC-AUC scores between 0.6051 and 0.6489 were obtained. Finally, we also make the datasets publicly available at https://robotics.ntua.gr/eprevention-sp-challenge/.

INDEX TERMS Person identification, relapse detection, anomaly detection, autoencoder architectures, biometric indexes, deep learning, digital phenotyping, long-term data, psychotic disorders, wearable technologies, ICASSP-2023 signal processing grand challenge.

I. INTRODUCTION

Nowadays wearable technologies offer unique opportunities to create innovative intelligent electronic services that can address various well-being issues. The widespread adoption of smartphones, as well as wearable products like smartwatches and fitness trackers, has given rise to the interdisciplinary field of digital phenotyping, which involves the quantification of human behavior and traits (the "phenotype") using sensors embedded in these devices. These

devices collect diverse data through geolocation sensors, accelerometers, gyroscopes, and heart rate monitors, measuring physical activity, kinetic activity, micro-movements, and autonomic function, opening thus the possibility for non-intrusive acquisition of activity, social and physiological data. This abundance of sensory data has facilitated the development of numerous applications for general user and health monitoring, as well as predictive analytic tasks such as emotional well-being [2], [3], sleep tracking [4], eating [5], agitation [6], and physical activity detection [7].

Over the past six decades, numerous studies have been conducted in the fields of neurobiology and neurophysiology to investigate psychotic conditions, including bipolar disorder and schizophrenia. Despite these efforts, the exact causes of these conditions remain elusive. As a result, no effective biomarkers for the prediction of psychotic symptoms have yet been discovered. Identification of early signs of worsening symptoms in the psychotic process and implementation of preventive measures have proven to significantly improve outcomes and mitigate the devastating impact that relapses can have on patients' lives [8], [9], [10]. Therefore, there is a growing emphasis on exploring the potential of such markers for timely diagnosis and prevention of psychotic relapses [11], [12], [13], [14], [15], [16], [17]. The current advances in the field of digital phenotyping offer the potential to support and revolutionize clinical psychiatry through the identification of biomarkers from passively collected sensory data and their correlation to the appearance of relapse episodes in patients, with the prospect of both transforming hospital-centered healthcare practice to proactive, individualized care, and improving the patient's course of life.

Several works have tried to tackle this problem offering promising evidence for using such sensory data from smartphones [14], [18], [19], [20] for characterizing the course of various mental illnesses, or identifying anomalies in periods before the appearance of relapses. The majority of these studies lasted up to a few weeks, with some exceptions [21]. Wearable sensors offer an unobtrusive and lightweight alternative for the monitoring of daily activities, since they are are deemed as comfortable by patients with mental disorders, and as such can be used as low-cost devices [16], [17], [22], [23] for the collection of physiological signals, such as accelerometer, gyroscope and heart rate measurements.

Supervised learning approaches for correlating the appearance of relapses with physiological data have mostly focused on either statistical significance testing or classification of hand-crafted features using traditional machine learning algorithms. Consequently, a variety of feature representations have been proposed in such medical settings, using data from wearables [24], [25]. Another approach used for relapse detection is non-supervised sensor-based anomaly detection, the importance of which has been highlighted during recent years and the COVID-19 pandemic, through the clinical mass adoption of telehealth [26]. This approach is especially suitable for mental health monitoring, where the availability of data corresponding to relapsing states is scarce, and has been applied on data collected from various passive sensors [1], [21], [27], [28], [29], [30]. Finally, with the increase in unlabeled amounts of data, self-supervised approaches are also emerging [31].

Regarding person identification from sensors embedded in smartwatches, kinetic data are mostly being used, in conjunction with simple statistical features extracted from the raw signals [32], [33], [34]. More recent approaches utilize deep learning, an approach that effectively generalizes in user-diverse datasets, while being able to disentangle the recorded signals and sensor noise with suitable data augmentation. In this case, CNN-based architectures are mostly being used, in either a discriminative fashion [35] or a representation learning framework [36].

However, the public availability of large user-diverse datasets of physiological signals is scarce, especially in conjunction with mental health indicators. As a result, through the e-Prevention Challenge,[1] researchers in the field had the opportunity to work on and draw insights from a large-scale collection of smartwatch signals (including continuous measurements from accelerometers, gyroscopes and heart rate monitors, as well as information about the daily step counts and sleep), collected from patients in the psychotic spectrum for a monitoring period of up to 2.5 years, and a control subgroup for a provisional period of 3 months, in two different tasks:

1) Studying the correlation of the biosignals to user-specific behavioral patterns via **Person Identification** from the recorded signals, and
2) Using these smartwatch signals as biomarkers of psychotic symptomatology for the task of **Relapse Detection** in psychotic patients.

Both tasks are of importance to the biomedical signal processing and psychiatry communities, since through the identification of digital phenotypes from wearable signals, useful insights on the distinctive behavioral patterns and relapse course of patients with mental disorders can be derived, contributing to early symptom identification, and eventually better outcomes of the disorder. Our teams have already undertaken extensive work on the full length e-Prevention dataset, pursuing the ultimate goal of relapse detection through person identification, either by addressing it as a misclassification problem [27] or employing a self-supervised framework combined with survival analysis for revealing the risk of an oncoming relapse [31]. Our research on relapse detection indicates that there is ample scope for further exploration, and the initial findings in this direction are promising. The ability to detect relapsing states from unobtrusive sensors is a first step towards the determination of biomarkers that correlate with the state of psychotic patients, and could

[1] See https://robotics.ntua.gr/eprevention-sp-challenge/ for the full challenge outcomes as well as the publicly available e-Prevention dataset for both tasks.

eventually lead to the prediction and finally the prevention of these relapses.

II. CHALLENGE OVERVIEW

A. DATASET

During the course of the e-Prevention project [1] (https://eprevention.gr/), a total of 60 people (37 patients in the psychotic spectrum and 23 healthy controls) were recruited at the University Mental Health, Neurosciences and Precision Medicine Research Institute "Costas Stefanis" (UMHRI) in Greece, and the protocol of the project was approved by the Ethics Committee of the Institution. All participants were provided with a Samsung Gear S3 smartwatch that monitored the user's linear acceleration and angular velocity (m/s^2 and deg/s, sampled at 20 Hz), heart rate variability and RR intervals (sampled at 5 Hz), sleeping schedule and steps.

This information was continuously collected from the patients for a monitoring period of up to 2.5 years, while the same data were collected from the control subgroup for a provisional period of 3 months. The resulting dataset is one of the largest of its kind ever recorded, with a total of approximately 20000 human-days of collected data spread among all participants. The collected data were anonymized, and each participant in the study was assigned a unique ID as an identifier. The clinicians annotated the patients' relapse periods according to their monthly assessments and communication with the attending physician or the family (more information about the recruitment and the monthly in-person clinical assessment can be found in [1]).

B. TASK DESCRIPTION

The e-Prevention challenge focused on two tasks, chosen as already mentioned for their significance, since the identification of digital phenotypes from wearable sensors provides valuable insights into the unique behavioral patterns and relapse course of patients with psychiatric disorders. In more detail, the two tasks were the following:

1) **Person Identification:** The goal in this task was to identify the watch wearer by forming and classifying their digital phenotypes from the recorded biosignals.
2) **Relapse Detection:** This task was aimed at detecting the appearance of relapses in the patients, based on the smartwatch measurements.

C. DATA FORMAT AND STRUCTURE

For the purposes of the e-Prevention Challenge, we provided two subsets of the collected dataset, one for each challenge task. For the Person Identification task, we provide a stratified split of a part of the dataset (including all 46 users, both patients and controls), consisting of about two and a half months per person. For the Relapse Detection task, the provided dataset constitutes again a subset of the full dataset, with data derived from a subgroup of ten patients corresponding approximately to six months per person.

For both tasks, we split the recordings of all users into days and shuffle them (i.e., there is no temporal continuity over different days). For each day, the data provided in both tasks include continuous signals of *linear acceleration* (from the accelerometer), *angular velocity* (from the gyroscope), *heart-rate* and *RR-interval* (from photoplethysmography - PPG). Before sharing the signals we aggregated the values of each signal over 5 seconds, resulting in 12 sample points in one minute. This was also done to mitigate the effect of each individual sensor noise to the classification task, which was observed in [35]. Apart from these signals, we also shared the sleep schedule of the users and walking information, along with the corresponding timestamps.

For the **Person Identification** task, each day in the training and validation split was also accompanied by the ID of the corresponding user - in the test set the teams were asked to predict the user ID. For the **Relapse Detection part** the training split contains only data acquired while the patient condition was stable, while the validation and testing splits span both stable and relapsing periods. The participants were expected to detect the relapses as anomalies and tackle this task with an unsupervised learning method, i.e., anomaly or novelty detection, using the labeled validation set to tune their algorithms. Finally, the testing set does not contain any information pertaining to the appearance of relapses.

D. BASELINES

For the baseline of the **Person Identification** task we trained a deep 1D CNN with 5 convolutional layers, including Batch Normalization and ReLU activations. After the last Batch Normalization layer we used adaptive average pooling and a final fully connected layer to predict the logits for the 46 identities. The data were preprocessed by normalizing the accelerometer, gyroscope and heart rate values into the [0–1] interval using their provided valid range, which for the tri-axial linear acceleration from the accelerometer was between [−19.6, 19,6], and for the tri-axial angular velocity from the gyroscope was between [−573, 573]. For the heart rate and the RR Intervals we set the max values as 255 and 2000 respectively.

During training, we sampled randomly 3-hour contiguous segments from the daily recordings, provided that they included at least 2.5 hours of valid data. Afterwards, any missing timestamps in the segment were imputed using Nearest Neighbor Interpolation (NNI), resulting in a sequence of 721 $\times$ 8 features (we did not use sleeping or step information). We fed the segment through the network, predicted the ID of the user and used Cross Entropy Loss for training. We trained for 300 epochs with a batch size of 64, using Adam with an initial learning rate of 1e-4 and reducing it by a factor of 10 at 150 and 225 epochs.

During inference, we selected all contiguous 3-hour segments of each daily recording with at least 1 h of valid data, imputed them again with nearest neighbors, and used voting over all segments in order to select the final predicted user

ID of the respective recording. Our method's final validation score was 62%.

The baseline provided for the **Relapse Detection** task is based on an 1-layer linear autoencoder. A total of 10 features were extracted from 5 minute slices of the original data. In more detail, the mean norm of the linear (accelerometer) and radial (gyroscope) measurements were computed to quantify movements and micro-movements, while cardiac behavior was estimated by the mean heart rate (bpm), the mean RR interval (ms), as well as the major axis of the Poincare ellipse and the normalized low- and high-frequency powers of the Lomb-Scargle periodogram as computed from the NNI series. Finally, daily sinusoidal encoding (sine and cosine values) was used for the temporal encoding of the timestamps, whereas the percentage of valid 5 s measurements was also calculated. Missing features were handled by median interpolation for intervals up to 3 hours; larger intervals of missing values were discarded. These features were then stacked into 2D tensors of size 48×10 (with each row representing a 5-minute slice and each column a feature), thus covering 4 hours each, using a stride of 1 h.

Then, an autoencoder was trained, with a bottleneck dimension equal to $N = 60$ and a LeakyReLU activation, using as input the feature representations mentioned above, after being standardized per-patient and flattened into a 480 dimensional vector. The post-normalization statistical properties (i.e., the mean and covariance matrix) of each (5-minute) feature slice in the training set were used to compute a multivariate normal distribution, that the feature vectors follow. During inference, input tensors, corresponding to 4 h intervals, are standardized according to the precomputed per-patient transform, and fed into the autoencoder. The per-feature mean of the autoencoder output is then computed, and its Mahalanobis distance to the assumed feature distribution is calculated and used as an anomaly score; input tensors corresponding to relapsing periods are expected to record higher anomaly scores. Since the evaluation is carried out in a per-day basis, the median anomaly score over all the 4 h tensors corresponding to each day was computed to obtain a single anomaly score for each day. Application of the above methodology in the provided validation set yielded a PR-AUC score of 0.635, a ROC-AUC score of 0.578 and a mean score of 0.6065.

E. EVALUATION METRICS

Person Identification: The proposed solution should return a prediction of the unique patient ID for daily intervals, by either aggregating predictions over smaller segments, or processing data corresponding to one day as a whole. Since person identification is a multi-class single-label problem, the weighted per-person identification accuracy is used as a metric.

Relapse Detection: Similarly to above, the evaluation of the state of the patient as stable or relapsing is carried out on a daily basis, either by suitable preprocessing of the input features or via post-hoc aggregation of predictions over smaller segments. Since this is an anomaly detection task, the mean of the PR-AUC (Precision-Recall Area Under Curve) and ROC-AUC (Receiver Operating Characteristic Area Under Curve) scores over the daily predictions is utilized as the main evaluation metric.

F. CHALLENGE RULES AND TIMELINE

1) RULES AND REQUIREMENTS

All participants should adhere to the following rules to be eligible for the challenge:

- All participants had to submit the obtained results for at least one of the 2 tasks, accompanied with a short (up to 1 page) description of their proposed system and methodology.
- Participating teams are allowed to update their submissions and scores multiple (up to 5) times during the evaluation phase.
- Each individual participant cannot be included in multiple participating teams.
- After the completion of the challenge, the top scoring teams for each track will be declared as winners of the respective track and they will be invited to provide a synopsis of their proposed methodology and results in a two-page-long paper, and present it in-person to the Special Session dedicated to the e-Prevention challenge in the ICASSP-2023 conference.
- Participants can only publish their own results, while a summary of the challenge results will also be prepared by the organizers.
- There are no restrictions on the proposed methodologies, as long as they follow the guidelines for each track, or the usage of external datasets. However, in case of a tie, the Challenge Committee will take into account the novelty and originality of the proposed approach.
- The intellectual property (IP) of all shared/submitted code, if applicable, remains to the participants and is not transferred to the challenge organizers. If the code developed by the participants is made publicly available, an appropriate license should be added.

2) DATA PERMISSION

Regarding the challenge data, permission was granted only if the participants agreed to the following terms:

1) To include a reference to the e-Prevention 2023 Dataset in any work that makes use of the dataset. For research papers, to cite the recommended publications (as listed on our website (https://robotics.ntua.gr/eprevention-sp-challenge/)) and the challenge overview paper.
2) To not distribute the dataset or modified versions.
3) To not use the dataset or any derivative work for commercial purposes as, for example, licensing or selling the data, or using the data with a purpose to procure a commercial gain.
4) All rights not expressly granted to the participants are reserved by the e-Prevention SP Grand Challenge 2023 organizers.

3) SUBMISSION INSTRUCTIONS

The winning teams' submitted paper had to be prepared according to the ICASSP-2023 guidelines (which we mention here in short for completeness). The 2 pages long paper, which should be as self-contained as possible, should follow the template for regular ICASSP papers including title, abstract, introduction, references and possible figures and tables. The abstract and introduction must clearly mention that the work was done in the context of an "ICASSP Signal Processing Grand Challenge" (including the official challenge name and the year of the challenge), and cite the selected publications. Additionally, the introduction should at least contain: a brief description of the scope of the challenge and a brief description of the participants' proposed solution and quantitative results obtained on the challenge's evaluation metrics. In the main text, the participating teams should focus on the conceptual implementation/innovation, and a high-level description of the proposed solution, while standard citation rules remained applicable, if as, for instance, the selected solution was inspired by (or used) existing work.

4) REGISTRATION AND SUPPORT

To register for the challenge, participants were required to send an e-mail to two organizing members (predefined), mentioning their team name, the names and emails of the team members, as well as their affiliations. The challenge participants were also encouraged to contact our team for any issue or clarification about the challenge or the dataset.

5) TIMELINE

- November 28th, 2022: Registration opens
- December 8th, 2022: Dataset Release and starting date
- February 1st, 2023: Deadline for participants to submit their results
- February 6th, 2023: Notification of the final results
- February 20th, 2023: Deadline for invited paper submission
- March 7th, 2023: ICASSP 2023 SPGC acceptance notification
- March 14th, 2023: ICASSP 2023 SPGC camera-ready papers

III. CHALLENGE RESULTS AND DISCUSSION

In total 15 teams successfully participated in the e-Prevention challenge and specifically 11 in the Person Identification task and 6 in the Relapse Detection task. Participating teams were allowed to compete in any, or both, tasks, and they could freely choose whether to work with data from all modalities, or specific ones (i.e., heart-rate information only). Two of the teams competed in both tasks, with one of those (PeRCeiVe) obtaining results that ranked in the top-3 submissions in both tasks. Table I shows the ranking with the final top-6 submissions, 3 for each task and the respective top-5 teams.

In general various methodologies were used by the different teams. Regarding the Person Identification task, the first stage included data preprocessing, such as normalization, data cleaning or even data augmentation. Six (6) out of 11 teams used deep architectures including 1D CNNs (Convolutional Neural Networks), combined 1D CNNs-LSTM (Long-Short-Term-Memory) networks, Transformers or ensemble methods based on the above architectures. Two teams used classical machine learning methods such as SVMs (Support Vector Machines) and GMMs (Gaussian Mixture Models), while the machine learning method utilized could not be reliably validated for the rest of the teams, as will be explained later. The top-3 ranked teams used deep architectures.

TABLE 1 Ranking With the Final top-6 Submissions, 3 for Each Task and the Respective Top-5 Teams.

Team ID	Person ID	Relapse Detection		
	Accuracy (%)	ROC-AUC	PR-AUC	Mean
SRCB-LUL [37]	**95.00%**	-	-	-
PeRCeiVe [38]	93.85%	**0.6469**	**0.6509**	**0.6489**
AI_Bezzie [39]	91.36%	-	-	-
Emotion [40]	-	0.6072	0.6347	0.6209
SAILers [41]	-	0.5839	0.6263	0.6051

The bold values correspond to the best value for each metric, denoting the best team of each track of the challenge, as well.

Continuing with the Relapse Detection task, the majority of the teams (5 out of 6 participating teams, including 2 out of the top-3 submissions) used autoencoder-based methods. 2 out of 5 teams used a Transformer-based autoencoder, 2 teams used 1-layer linear autoencoders - in one case, the autoencoder's output was post-processed by an LSTM for temporal feature aggregation, and 1 experimented with a CNN-based autoencoder, a U-Net and a Transformer Encoder-Decoder, yielding the best results for this task. Notably, the only team not using an autoencoder-based method (team SAILers), which utilized a tree-based ensemble method, ranked in the top-3 teams of the track. Similar to the previous task, all teams followed a preprocessing procedure including normalization and feature extraction.

An overview of the proposed methodologies can be found in Tables II and III for the person identification task and in Tables IV and V for the relapse detection task. Note that while it was required by the participants to send an 1-page description of their proposed methodologies along with their results, that was not the case for all teams that participated in task 1; that is why the respective tables present the solutions of 7 out of 11 participating teams only. Finally, the top-6 submissions by the 5-top ranked teams are discussed in more detail in the following section.

A. METHODOLOGY DESCRIPTION OF THE TOP-SIX SUBMISSIONS

1) PERSON IDENTIFICATION TASK

In [37] team SRCB-LUL developed a system for the person identification task, taking the first ranking place with an accuracy of 95%. They used eight-channel signals, including accelerometer, gyroscope, heart rate, and RR intervals, while

TABLE 2 Person Identification (Task 1) Methodologies Used Regarding the Preprocessing Steps Followed by the Participants and Specifically, the Missing Value Handling, the Processing Segment Resolution and the Normalization. The Score (Accuracy %) Obtained by the Various Teams is Also Presented.

Team ID	Score	Missing Value Handling	Processing Segment Resolution	Normalization
SRCB-LUL [37]	95.00 %	Mean replacement over day	30 min, 5 sec slices	Mean-Std norm
PeRCeiVe [38]	93.85 %	NNI, discard if many invalid data	1.5 hour, or 3 hours (5 sec slices)	Min-Max norm [0,1]
AI_Bezzie [39]	91.36 %	Development of missingness-aware framework	1 hour, 30 sec slices	Yes (type not stated)
SAILers	82.15 %	Discard missing data	Daily, 3 min slices	NS
Unipi-CMBL	75.43 %	NS	Daily, 5 sec slices	Min-Max norm [0,255]
CogBCI	2.88 %	NS	1 hour, 5 sec slices	NS
UOI	2.68 %	Discard missing/invalid data	3 hours, 30 min slices	None

(NS denotes not stated.)

TABLE 3 Person Identification (Task 1): Features, Modalities, and Model(s) Used by the Participating Teams.

Team ID	Acc	Gyr	HRV Feats	Sleep	Step	Feat Aggregation	Model	Ensemble
SRCB-LUL [37]	3-axes	3-axes	heart rate, RR	yes	no	no	1D-CNNs (mid-depth)	yes
PeRCeiVe [38]	3-axes	3-axes	heart rate, RR	yes	yes	no	Transformers /1D-CNN fusion	yes
AI_Bezzie [39]	3-axes	3-axes	heart rate, RR	yes	yes	no	1D-CNN+LSTM	no
SAILers	norm	norm	heart rate, RR	yes	yes	yes (mean+std)	GMMs (late modality fusion)	yes
Unipi-CMBL	3-axes	3-axes	heart rate, RR	yes	no	no	2D-CNN (Imagenet Pretraining)	no
CogBCI	3-axes	3-axes	heart rate, RR	no	no	no	CNN+LSTM/CNN fusion	yes
UOI	3-axes	3-axes	heart rate, RR	no	no	yes (extraction of statistical features from all raw signals + entropies)	DNN (3 layers)	no

the step information was discarded. The outlier values were filtered out and replaced by the mean value of the day, while the data were also normalized according to the distribution of the training data. The valid data were divided into multiple short-term segments, in order to solve the problem of abnormal values, which were then used to predict the identification results. Afterwards, multiple base classifiers were trained, by changing the segment length and the number of the signals, taking into account the wakefulness state (awake vs. asleep), and an ensemble model was used to obtain the final results of the user ID. For the training, they designed a 1D CNN with multiple convolutional layers according to [42], which were followed by a batch normalization layer and a ReLU activation. After concatenating the intermediate features multiple fully connected (FC) layers followed. Except for the last layer, each fully connected layer was followed by a ReLU activation. Finally, the logits for the 46 identities were output. During inference, the predicted user ID was obtained by voting over both sleep and awake segments.

In [38] for the task of person identification the PeRCeiVe team, which ranked second with an accuracy of 93.85%, first normalized the data, including accelerometer, gyroscope, heart rate, and RR intervals within the range of [0, 1], retaining the sleep information, while invalid data were imputed with nearest neighbor interpolation. Three sequences were derived from step data: the number of steps per time unit, velocity in m/s, and calories per time unit. To augment the data, a sliding window approach with non-overlapping windows of varying widths (1.5 and 3 hours) was employed, resulting in input sequences of $T \times F$, where T denotes the resolution of each temporal slice and F corresponds to the 12 series data, including 3 accelerometer, 3 gyroscope, heart rate, RR intervals, sleeping activity, and the three step series. They implemented an ensemble model comprising a standard deep 1D CNN and five transformer architectures. The CNN model also served as the embedding backbone for all the transformer models. Time2Vec [43] was used as positional embedding, and a [CLS] token was added to the final sequence. Different transformer configurations, involving positional encoding, model depth, encoding layers, and attention heads, were utilized in the ensemble. During training, only windows with at least 83% valid data were considered, with invalid data imputed using NN interpolation. The optimal parameters for each model were chosen based on the performance on the

TABLE 4 Relapse Detection (Task 2) Methodologies Used Regarding the Preprocessing Steps Followed by the Participants and Specifically, the Missing Value Handling, the Processing Segment Resolution and the Normalization. The Score (Mean PR- and ROC-AUC) Obtained by the Various Teams is Also Presented.

Team ID	Score	Missing Value Handling	Processing Segment Resolution	Normalization
PeRCeiVe [38]	0.6489	NNI, discarding if many invalid data	1.5 hour, 5 sec slices (raw features)	Min-Max norm [0,1]
		NNI, discarding if many invalid data	4-hours 5 min slices (aggregated features)	Per-Patient Mean+Std norm
Emotion [40]	0.6209	Median per-patient replacement	1 day, 4 hour slices	Per-Patient Mean+Std norm, followed by MinMax across all patients
SAILers [41]	0.6051	Remove outliers; Hampel interpolation for small intervals, discard large ones	Daily (sleep-filter), 5 min slices	Unit-norm for each feature per patient
SmartBCI	0.5604	NS	4 hours, unknown slice length	Mean+Std norm
YDH@HEU	0.5401	NS	4 hours, 5 min slices	NS
GISP@HEU	0.5229	NS	4 hours, 5 min slices	NS

(NS denotes not stated.)

TABLE 5 Relapse Detection (Task 2): Features, Modalities, Models, and Anomaly Measure Used by the Participating Teams.

Team ID	Acc	Gyr	HRV Feats	Sleep	Step	Time	Model	Anomaly Measure	PM
PeRCeiVe [38]	3-axes (raw)	3-axes (raw)	heart rate, RR	yes	yes	no	1D-CNNs autoencoders	reconstruction error distribution likelihood (CDF)	yes
	mean norm	mean norm	mean heart rate & RR, Poincare major axis, Lomb-Scargle LF, HF	no	no	yes	1D-CNN/Transformer based autoencoders	reconstruction error distribution likelihood (CDF)	yes
Emotion [40]	mean norm	mean norm	Mean + Std. rate	yes	yes	no	Linear autoencoder (10 bottleneck features)	reconstruction error	no
SAILers [41]	mean norm	mean norm	mean heart rate, HRV, Welch LF, HF (+ fractions)	yes	yes	yes	Isolation Forest	number of splits required for sample isolation	no
SmartBCI	mean norm	mean norm	mean heart rate & RR, HRV, Lomb-scargle VLF, LF, HF (+ ratios)	no	no	no	Linear autoencoder (60 bottleneck features)	NS	no
YDH@HEU	mean norm	mean norm	mean heart rate & RR	no	no	yes	Transformer autoencoder	NS	no
GISP@HEU	mean norm	mean norm	mean heart rate & RR	no	no	yes	Conformer autoencoder	NS	no

NS denotes not stated by the authors, while PM denotes personalized models.

validation set, and inference prediction was performed by summing the prediction logits of each model in the ensemble.

Finally, for the same task the AI_Bezzie team [39] that ranked third with 91.36% accuracy resampled all data with a 30 s interval and used a union mechanism to extract intersecting time periods. They categorized the time of the day into four segments and used one-hot-encoding to represent it as a static feature. Additionally, a missing feature encoding was used to indicate data loss in the heart rate sensor. All data streams were normalized based on specific sensors' characteristics. Their architecture utilized a fixed window size with a 30% overlapping ratio for data slicing during training. Inference involved predicting labels for all available slices in a day and using majority voting to assign the final label. Two types of feature fusion were also performed: (a) early fusion for the temporal features and (b) late fusion for the static features. Early fusion used individual feature encoders [35] using three 1D-Convolutional layers, while the static features (i.e., sleep information and time of the day among others) were combined with the output of the sequence model before being fed to the next layer. A single-layer LSTM was utilized to model the temporal relationship between the data and Adam optimizer was used for the 46-class classification problem with Cross-Entropy loss, while various ablations studies were conducted regarding the window length, the features used and the modifications in the architectures.

2) RELAPSE DETECTION TASK

Continuing with the second task of the e-Prevention challenge, PeRCeiVe team [38] participated in the relapse detection task, as well. In this task, they ranked first with a mean ROC-AUC and PR-AUC score of 0.6489. An anomaly detection in a personalized scheme approach was followed utilizing both raw and aggregated data, where raw data were processed similarly to the person identification task, while the aggregated approach involved extracting 10 features from 5-minute slices of the original data. The features included the mean norm of linear accelerations and radial velocities, mean heart rate and RR interval, the major axis of the Poincare ellipse, normalized low and high-frequency powers of the Lomb-Scargle periodogram from the NNI series, daily sinusoidal encoding, and percentage of valid samples. Data were standardized using the per-patient mean and standard deviation as computed on the training set. Three different architectures –a CNN-based autoencoder, an autoencoder for time series, and a Transformer Encoder-Decoder– were employed to create patient-specific models. These models were trained separately on both raw and aggregated data using non-relapse data and the Mean Squared Error (MSE) as a loss function. The model selection was based on the best accuracy on the validation set. During inference, the reconstruction error was computed, and the Cumulative Distribution Function (CDF) of the per-channel reconstruction error was used as an anomaly score. The evaluation was carried out using the median anomaly score over all available windows for each day.

Emotion team [40] that ranked second with a mean ROC-AUC and PR-AUC score of 0.6209, for the task of relapse detection, performed data cleaning before feature extraction, i.e., duplicate time intervals were removed and heart rate values outside a specific range were discarded. Additionally, heart rate values deviating more than 20% from the heart rate calculated using RR intervals were eliminated. Negative step counts in the physical activity data were also removed. Missing values were replaced with the median of each feature for each patient. Features were then computed for each 4-hour period, including mean and standard deviation of the heart rate, norm of accelerometer and gyroscope coordinates, percentage of sleeping time, and total number of steps. These features were concatenated into a single vector for each day. Standard normalization was applied to each patient's data, followed by min-max normalization across all patients' data. To predict relapse in patients with psychotic disorders, a simple auto-encoder (AE) neural network with 1 hidden layer and 10 neurons was implemented (other anomaly detection techniques were also evaluated, however showing that they perform similarly, with the AE achieving the highest overall score). Their autoencoder architecture was trained using the Adam optimizer and MSE as the loss function, while the averaged reconstruction error was utilized as an anomaly score to detect relapses in patients.

Finally, for the task of relapse detection in [41] team SAILers, who ranked third with a mean ROC-AUC and PR-AUC score of 0.6051, removed outliers and applied the Hampel method [44] to impute missing values within a 1-hour range. Feature extraction involved deriving various features from 5 m intervals of the processed time-series, which were then aggregated at different resolutions. In more detail, the normalized energy of accelerometer and gyroscope data was calculated from the tri-axial measurements, the mean heart rate and heart rate variability (HRV) were extracted from RR intervals, and their power spectral density was estimated using Welch's method to isolate low- and high-frequency bands, along with their respective fractions. Timestamps were encoded using daily sinusoidal encoding. Regarding the step count data, they integrated the provided features and calculated additional step size and speed by converting the start and end times of steps into seconds. The data was then distributed over 5 m intervals by summing up the step counts and taking the mean of distance, calories, step size, and speed. After feature extraction for each patient, they standardized the features to unit norm and concatenated them in a daily basis to create subject-agnostic trials, approaching the problem as a novelty detection task, where the goal was to identify outliers without any outlier data in the training set. Towards this end, they selected the Isolation Forest [45], a tree-based ensemble method, where features were randomly selected and split between extreme values. The number of splits required to isolate a sample served as a measure of normality, since anomalies are likely to have shorter paths due to random partitioning. This measure is averaged over a forest of such random trees to assess outlier detection performance. In their work, the use of sleep activity, step count and heart rate features were also investigated evaluating various combinations of features and time resolutions.

B. DISCUSSION: TRENDS, CHALLENGES AND ADVANCEMENTS

Some of the key trends across the 3-top submissions regarding the Person Identification task included addressing abnormal sensor data through segmentation, thus creating segments varying in length in order to solve the problem of both abnormal and missing values, while generally using voting over all segments in order to obtain the final prediction results. Additionally, the different teams, apart from the main eight-channel signals (including accelerometer (3-axis), gyroscope (3-axis), heart rate and RR intervals), leveraged the additional signal modalities. Regarding the step information provided, some used all provided signals, even the information of calories [39], while others discarded completely the step counts [37]; on the other hand, all teams retained the wake vs. asleep information. Moreover, they employed ensemble models (either unimodal or multimodal), and utilized normalization and various imputation techniques, showcasing the significance of preprocessing, model architecture modifications and hyperparameter tuning to achieve higher accuracy. In addition, a necessary "good practice", since it was adopted by all three leading teams, involves addressing missing data through their

replacement or some kind of "coding", rather than discarding them outright. Among the proposed methodologies the introduction of a missing feature encoding method to indicate data loss showed promising results compared to statistical imputation techniques [39]. Moreover, as also observed in Table III, 1D CNN-based architectures constitute a part of all top-3 submissions. The focus on temporal and static feature fusion, such as early and late fusion strategies, highlighted the importance of optimizing the feature integration. Finally, through ablation studies that were conducted, one team [39] concluded that in their case the kinetic features from the accelerometer contained the most discriminative features, while the modality of gyroscope had the least impact. This behavior of the gyroscope-derived features was also observed in [31], for the task of relapse prediction.

Continuing with the second task of Relapse Detection, some of the trends among the 3-top submissions include the use of anomaly detection methodologies, leveraging various physiological signal features, and employing diverse machine learning techniques such as autoencoders [38], [40] and in one case a tree-based ensemble method [41] (see also Tables IV and V). The 2-top submissions [38], [40] utilized autoencoders, indicating that they can achieve better results in the task. The various strategies focused on data preprocessing and feature engineering, demonstrating the importance of addressing missing values, abnormal data, and – with a larger emphasis compared to the person identification task, and some variability among the top teams – identifying the most informative features, or even modalities across the ones provided in the e-Prevention dataset. In addition, we hypothesize that deep and complex architectures require targeted pre-processing and normalization procedures in order to unlock their full potential, since as can be also seen by the results in Tables I and IV, even the non-deep tree-based Isolation Forest method used in [41] obtains rather good results, outperforming submissions that utilized Transformer-based architectures without targeted pre-processing. In this task as well, the various teams used different segment lengths for their experimentation, however compared to the person identification task the temporal windows used are larger, with two out of three top teams using a daily resolution. Furthermore, in [41], the authors investigated the use of sleep activity, step count and heart rate features and evaluated various combinations of features and time resolutions, showing that short-time sleep behavior features outperformed their awake counterparts as well as the larger time intervals. Finally, a valuable practice for good performance in the relapse detection task is the use of some degree of personalization strategies, as also seen in [14], [28], either by using outright personalized models or per-patient normalization techniques, while utilizing global models.

The results of the challenge indicate an advancement of the state-of-the-art in both the person identification task and the psychotic relapse detection task. With regards to the person identification task, novel solutions have been developed for the handling of missing features, such as employing a separate encoding for replacing them [39], whereas ensemble models, trained in separate modalities and/or with diverse training settings, yielded better results than the commonly employed end-to-end models [35], [36]. The main advancement, however, concerns the swing at the direction of raw sensorial data instead of aggregated feature representations; the top-performing teams either utilized directly the provided data at the 5-sec resolution [37], [38], or downsampled them, with the final sampling rate remaining less than 1 minute [39]. This contrasts with the state-of-the-art in the task, where mostly spectral features [36] have been employed, and a larger (minute-scale) temporal slice has been used for feature aggregation [27], [31] (with a few exceptions [46]), and implies the potential of achieving further gains with the utilization of raw recordings. On the other hand, for the relapse detection task, the main advancement concerns successful adaptation of Transformer-based architectures [38], in contrast to DNN or CNN ones that reached the best performance in previous works [21], [28]. Another direction that, based on the results of the challenge, merits further exploration regards emphasizing the behavior of the patients while asleep; features extracted from biosignals acquired during sleep resulted in higher relapse detection rates compared to their awake counterparts [41], which is in agreement with psychiatric findings in the cases of both bipolar disorder [47] and schizophrenia [48].

Finally, regarding the scores obtained in the two distinct tracks (see also Table I); in the person identification task, we may see that the 3-top teams, all utilizing deep architectures, obtained accuracy results between 91%–95%. The next best submissions, with accuracy results between 75%–83%, utilized either deep or traditional machine learning methods, while a number of submissions performed close to random chance for the task. This both highlights the superiority of deep learning based methodologies for the task and the necessity for efficient data preprocessing schemes, as described previously. Regarding the relapse detection task, the results, which are over or comparable to the baseline (with mean PR- and ROC-AUC scores between 0.6051 and 0.6489), showed that the specific task is indeed a difficult task to solve, especially in comparison to the (multi-label) person identification task. Since nowadays the detection or even the prediction of relapses is considered of major importance in the field of psychiatry, the results imply that more sophisticated data preprocessing schemes, more powerful network architectures, additional modalities, or external information such as demographics [31] should be explored towards this goal.

IV. CONCLUSION

In this paper, we provide an overview of ICASSP-2023 e-Prevention challenge, conducted with the goal of analyzing the ability of digital phenotypes in quantifying behavioral patterns tackling two different but very important tasks; the task of Person Identification and Relapse Detection. To this end, long-term continuous recordings of the e-Prevention project, collected from unobtrusive smartwatches, were provided to the participants. The top participating approaches, for both

tasks, yielded notable performance, acquiring an accuracy of 95% in the Person Identification task and a mean PR- and ROC-AUC score of 0.6489 in the Relapse Detection task. The participants explored a variety of methodologies, ways of handling missing and abnormal data, while other team-specific approaches, regarding either the selection of the modalities, or the length of the segments or the specific architectures used, contributed to this boost in performance.

Overall, we believe that the e-Prevention dataset holds a lot of promise for hosting future e-Prevention challenges and incentivising further research. As a result, we plan to organize more challenges with additional tracks that include depressive relapses, apart from psychotic ones. Two limitations of the presented challenge was the fact that the shared signals were aggregated over 5 seconds and that there was no temporal continuity over different days –in the future we plan to both release raw data and share larger sequences that will allow for architectures that can leverage large temporal windows. Finally, apart from sleeping and wakefulness information it would be interesting to see if there are any correlations behind our collected demographic data of the patients (which were not given at this challenge) and see if they can be leveraged to predict relapses.

In this direction, we are currently organizing the 2nd e-Prevention challenge: *Psychotic and Non-Psychotic Relapse Detection using Wearable-Based Digital Phenotyping*[2] for ICASSP-2024. The objective is to stimulate innovative research across two distinct tasks: 1) Detection of non-psychotic relapses, and 2) Detection of psychotic relapses, both in patients within the psychotic spectrum; hereby pushing boundaries in this important area of mental health care.

IV. NOTE ON ETHICAL CONSIDERATIONS AND PRIVACY

While the development of personalized and adaptive healthcare has benefited from the upsurge in the amount of passively collected and transmitted data from non-invasive sensors, numerous concerns are being raised about both the confidentiality of the collected data and the privacy of the participants in such studies [49]. This is especially true for the case of person identification, where systems trained to identify specific individuals can have potentially malicious use cases. In the e-Prevention project personal information about the participants is protected through assigning a unique ID to each participant, with the correspondence available only to the clinical team, as well as using a secure cloud server for data storage. Additionally, the person identification problem is faced as a classification problem of the signals to each participant's encoded ID, while it does not expose any further information about the users' identity. In general, researchers in the field should apply strict data security and privacy protocols, while at the same time apply data anonymization algorithms [50] when applicable, in order to mitigate those risks.

[2]More information about the 2nd e-Prevention challenge can be found at: https://robotics.ntua.gr/icassp2024-eprevention-spgc/ and https://2024.ieeeicassp.org/sp-grand-challenges/

ACKNOWLEDGMENT

The authors would like to thank all volunteers, controls and patients with psychotic disorders that participated in this long term study and provided the data; without them this project would not be possible. Finally, authors would like to thank the anonymous reviewers for all their suggestions for improving this paper.

INSTITUTIONAL REVIEW BOARD STATEMENT

All subjects gave their written informed consent and permission for inclusion and use of their anonymized data before they participated in the study.

REFERENCES

[1] A. Zlatintsi et al., "E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures," *Sensors*, vol. 22, no. 19, 2022, Art. no. 7544.

[2] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interaction*, 2018, pp. 400–408.

[3] B. Ballinger et al., "DeepHeart: Semi-supervised sequence learning for cardiovascular risk prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.

[4] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.

[5] E. Thomaz, I. Essa, and G. D. Abowd, "A practical approach for recognizing eating moments with wrist-mounted inertial sensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 1029–1040.

[6] R. Alam et al., "Motion biomarkers for early detection of dementia-related agitation," in *Proc. 1st Workshop Digit. Biomarkers*, 2017, pp. 15–20.

[7] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," in *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 3, pp. 1–30, 2019.

[8] T. R. Insel, "The arrival of preemptive psychiatry," *Early Interv. Psychiatry*, vol. 1, no. 1, pp. 5–6, 2007.

[9] M. Bertelsen et al., "Five-year follow-up of a randomized multicenter trial of intensive early intervention vs standard treatment for patients with a first episode of psychotic illness: The OPUS trial," *Arch. Gen. Psychiatry*, vol. 65, no. 7, pp. 762–771, 2008.

[10] W. t. V. Hegelstad et al., "Long-term follow-up of the TIPS early detection in psychosis study: Effects on 10-year outcome," *Amer. J. Psychiatry*, vol. 169, no. 4, pp. 374–380, 2012.

[11] D. Wiersma, F. J. Nienhuis, C. J. Slooff, and R. Giel, "Prodromes and precursors: Epidemiologic data for primary prevention of disorders with slow onset," *Amer. J. Psychiatry*, vol. 152, 1995, Art. no. 967.

[12] N. Koutsouleris et al., "Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification," *Schizophrenia Bull.*, vol. 38, pp. 1200–1215, 2011.

[13] P. McGorry et al., "Biomarkers and clinical staging in psychiatry," *World Psychiatry*, vol. 13, pp. 211–223, 2014.

[14] B. Lamichhane, J. Zhou, and A. Sano, "Psychotic relapse prediction in schizophrenia patients using a personalized mobile sensing-based supervised deep learning model," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 7, pp. 3246–3257, Jul. 2023.

[15] N. Long et al., "A scoping review on monitoring mental health using smart wearable devices," *Math. Biosci. Eng*, vol. 19, no. 8, pp. 7899–7919, 2022.

[16] S. Spulber, F. Elberling, J. Svensson, M. Tiger, S. Ceccatelli, and J. Lundberg, "Patterns of activity correlate with symptom severity in major depressive disorder patients," *Transl. Psychiatry*, vol. 12, no. 1, 2022, Art. no. 226.

[17] M. Razavi, A. McDonald, R. Mehta, and F. Sasangohar, "Evaluating mental stress among college students using heart rate and hand acceleration data collected from wearable sensors," 2023, *arXiv:2309.11097*.
[18] A. Maxhuni, A. Muñoz-Meléndez, V. Osmani, H. Perez, O. Mayora, and E. F. Morales, "Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients," *Pervasive Mobile Comput.*, vol. 31, pp. 50–66, 2016.
[19] J. J. Chapman, J. A. Roberts, V. T. Nguyen, and M. Breakspear, "Quantification of free-living activity patterns using accelerometry in adults with mental illness," *Sci. Rep.*, vol. 7, 2017, Art. no. 43174.
[20] I. Barnett, J. Torous, P. Staples, L. Sandoval, M. Keshavan, and J.-P. Onnela, "Relapse prediction in schizophrenia through digital phenotyping: A pilot study," *Neuropsychopharmacol.*, vol. 43, pp. 1660–1666, 2018.
[21] D. A. Adler et al., "Predicting early warning signs of psychotic relapse from passive sensing data: An approach using encoder-decoder neural networks," *JMIR mHealth uHealth*, vol. 8, no. 8, 2020, Art. no. e19962.
[22] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, vol. 3, pp. 1321–1330, Mar. 2015.
[23] D. Robotham, S. Satkunanathan, L. Doughty, and T. Wykes, "Do we still have a digital divide in mental health? A five-year survey follow-up," *J. Med. Internet Res.*, vol. 18, 2016, Art. no. e309.
[24] A. Voss, V. Baier, S. Schulz, and K. Bar, "Linear and nonlinear methods for analyses of cardiovascular variability in bipolar disorders," *Bipolar Disord.*, vol. 8, pp. 441–452, 2006.
[25] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, "Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches," *IEEE Access*, vol. 8, pp. 38 146–38 163, 2020.
[26] V. Canady, "Psychiatrists, patients favor telemedicine; challenges exist," *Ment. Health Weekly*, vol. 30, no. 20, pp. 4–5, 2020.
[27] N. Efthymiou et al., "From digital phenotype identification to detection of psychotic relapses," in *Proc. IEEE Int. Conf. Healthcare Informat.*, Houston, TX, USA, 2023, pp. 276–284.
[28] M. Panagiotou, A. Zlatintsi, P. P. Filntisis, A. J. Roumeliotis, N. Efthymiou, and P. Maragos, "A comparative study of autoencoder architectures for mental health analysis using wearable sensors data," in *Proc. 30th Eur. Signal Process. Conf.*, Belgrade, Serbia, 2022, pp. 1258–1262.
[29] K. Wang et al., "Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals," *Sci. Program.*, vol., 2016, 2016, Art. no. 5642856.
[30] J. Zhou et al., "Predicting psychotic relapse in schizophrenia with mobile sensor data: Routine cluster analysis," *JMIR mHealth uHealth*, vol. 10, no. 4, 2022, Art. no. e31006.
[31] E. Fekas, A. Zlatintsi, P. Filntisis, C. Garoufis, N. Efthymiou, and P. Maragos, "Relapse prediction from long-term wearable data using self-supervised learning and survival analysis," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–5.
[32] J. Ranjan and K. Whitehouse, "Object hallmarks: Identifying object users using wearable wrist sensors," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 51–61.
[33] W. U. Rehman, A. Laghari, and Z. Memon, "Exploiting smart phone accelerometer as a personal identification mechanism," *Mehran Univ. Res. J. Eng. Technol.*, vol. 34, 2015, Art. no. S1.
[34] A. Buriro, R. Van Acker, B. Crispo, and A. Mahboob, "AirSign: A gesture-based smartwatch user authentication," in *Proc. Int. Carnahan Conf. Secur. Technol.*, 2018, pp. 1–5.
[35] G. Retsinas, P. P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi, and P. Maragos, "Person identification using deep convolutional neural networks on short-term signals from wearable sensors," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 3657–3661.
[36] E. Maiorana, C. Romano, E. Schena, and C. Massaroni, "BIOWISH: Biometric recognition using wearable inertial sensors detecting heart activity," *IEEE Trans. Dependable Secure Comput.*, vol. 21, no. 2, pp. 987–1000, Mar./Apr. 2024, doi: 10.1109/TDSC.2023.3268360.
[37] J. Wu and M. Tu, "A person identification system for the ICASSP 2023 e-prevention challenge," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–2.
[38] S. Calcagno, R. Mineo, D. Giordano, and C. Spampinato, "Ensemble and personalized transformer models for subject identification and relapse detection in e-prevention challenge," in *Proc. Int. Conf Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–2.
[39] P. Mohapatra, A. Pandey, S. Keten, W. Chen, and Q. Zhu, "Person identification with wearable sensing using missing feature encoding and multi-stage modality fusion," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–2.
[40] S. Hamieh, V. Heiries, H. Al Osman, and C. Godin, "Relapse detection in patients with psychotic disorders using unsupervised learning on smartwatch signals," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–2.
[41] K. Avramidis, K. Adsul, D. Bose, and S. Narayanan, "Signal processing grand challenge 2023-e-prevention: Sleep behavior as an indicator of relapses in psychotic patients," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Rhodes Island, Greece, 2023, pp. 1–2.
[42] Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ecg signals," *Comput. Biol. Med.*, vol. 102, pp. 411–420, 2018.
[43] S. M. Kazemi et al., "Time2vec: Learning a vector representation of time," 2019, *arXiv:1907.05321*.
[44] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1635–1647, 2004.
[45] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 413–422.
[46] L. Alawneh, M. Al-Zinati, and M. Al-Ayyoub, "User identification using deep learning and human activity mobile sensor data," *Int. J. Inf. Secur.*, vol. 22, no. 1, pp. 289–301, 2023.
[47] A. K. Gold and L. G. Sylvia, "The role of sleep in bipolar disorder," *Nature Sci. Sleep*, vol. 8, pp. 207–214, 2016.
[48] D. S. Manoach, J. Q. Pan, S. M. Purcell, and R. Stickgold, "Reduced sleep spindles in schizophrenia: A treatable endophenotype that links risk genes to impaired cognition," *Biol. Psychiatry*, vol. 80, no. 8, pp. 599–608, 2016.
[49] J. Tu and W. Gao, "Ethical considerations of wearable technologies in human research," *Adv. Healthcare Mater.*, vol. 10, no. 17, 2021, Art. no. 2100127.
[50] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Mobile sensor data anonymization," in *Proc. Int. Conf. Internet Things Des. Implementation*, 2019, pp. 49–58.